



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Playing Anonymous Games using Simple Strategies

Citation for published version:

Cheng, Y, Diakonikolas, I & Stewart, A 2017, Playing Anonymous Games using Simple Strategies. in *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, pp. 616-631, Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, Barcelona, Spain, 16/01/17. <https://doi.org/10.1137/1.9781611974782.40>

Digital Object Identifier (DOI):

[10.1137/1.9781611974782.40](https://doi.org/10.1137/1.9781611974782.40)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Playing Anonymous Games using Simple Strategies

Yu Cheng*

Ilias Diakonikolas†

Alistair Stewart‡

Abstract

We investigate the complexity of computing approximate Nash equilibria in anonymous games. Our main algorithmic result is the following: For any n -player anonymous game with a bounded number of strategies and any constant $\delta > 0$, an $O(1/n^{1-\delta})$ -approximate Nash equilibrium can be computed in polynomial time. Complementing this positive result, we show that if there exists any constant $\delta > 0$ such that an $O(1/n^{1+\delta})$ -approximate equilibrium can be computed in polynomial time, then there is a fully polynomial-time approximation scheme (FPTAS) for this problem.

We also present a faster algorithm that, for any n -player k -strategy anonymous game, runs in time $\tilde{O}((n+k)kn^k)$ and computes an $\tilde{O}(n^{-1/3}k^{11/3})$ -approximate equilibrium. This algorithm follows from the existence of simple approximate equilibria of anonymous games, where each player plays one strategy with probability $1 - \delta$, for some small δ , and plays uniformly at random with probability δ .

Our approach exploits the connection between Nash equilibria in anonymous games and Poisson multinomial distributions (PMDs). Specifically, we prove a new probabilistic lemma establishing the following: Two PMDs, with large variance in each direction, whose first few moments are approximately matching are close in total variation distance. Our structural result strengthens previous work by providing a smooth tradeoff between the variance bound and the number of matching moments.

1 Introduction

Anonymous games are multiplayer games in which the utility of each player depends on her own strategy, as well as the number (as opposed to the identity) of other players who play each of the strategies. Anonymous games comprise an important class of succinct games

— well-studied in the economics literature (see, e.g., [Mil96, Blo99, Blo05]) — capturing a wide range of phenomena that frequently arise in practice, including congestion games, voting systems, and auctions.

In recent years, anonymous games have attracted significant attention in TCS [DP07, DP08, DP09, DP15, GT15, CDO15, DDKT16, DKS16a], with a focus on understanding the computational complexity of their (approximate) Nash equilibria. Consider the family of anonymous games where the number of players, n , is large and the number of strategies, k , is bounded. It was recently shown by Chen *et al.* [CDO15] that computing an ϵ -approximate Nash equilibrium of these games is PPAD-Complete when ϵ is exponentially small, even for anonymous games with 5 strategies¹.

On the algorithmic side, Daskalakis and Papadimitriou [DP07, DP08] presented the first polynomial-time approximation scheme (PTAS) for this problem with running time $n^{(1/\epsilon)^{\Omega(k)}}$. For the case of 2-strategies, this bound was improved [DP09, DDS12, DP15] to $\text{poly}(n) \cdot (1/\epsilon)^{O(\log^2(1/\epsilon))}$, and subsequently sharpened to $\text{poly}(n) \cdot (1/\epsilon)^{O(\log(1/\epsilon))}$ in [DKS16b].

In recent work, Daskalakis *et al.* [DDKT16] and Diakonikolas *et al.* [DKS16a] generalized the aforementioned results [DP15, DKS16b] to any fixed number k of strategies, obtaining algorithms for computing ϵ -well-supported equilibria with runtime of the form $n^{\text{poly}(k)} \cdot (1/\epsilon)^{k \log(1/\epsilon)^{O(k)}}$. That is, the problem of computing approximate Nash equilibria in anonymous games with a fixed number of strategies admits an *efficient* polynomial-time approximation scheme (EPTAS). Moreover, the dependence of the running time on the parameter $1/\epsilon$ is *quasi-polynomial* — as opposed to exponential.

We note that all the aforementioned algorithmic results are obtained by exploiting a connection between Nash equilibria in anonymous games and Poisson multinomial distributions (PMDs). This connection — formalized in [DP07, DP08] — translates constructive upper bounds on ϵ -covers for PMDs to upper bounds on com-

*Department of Computer Science, University of Southern California. Email: yu.cheng.1@usc.edu. Supported in part by Shang-Hua Teng’s Simons Investigator Award.

†Department of Computer Science, University of Southern California. Email: diakonik@usc.edu. Supported by a USC startup fund.

‡Department of Computer Science, University of Southern California. Email: alistais@usc.edu.

¹[CDO15] showed that computing an equilibrium of 7-strategy anonymous games is PPAD-Complete, but 3 of the 7 strategies in their construction can be merged, resulting in a 5-strategy anonymous game.

puting ϵ -Nash equilibria in anonymous games (see Section 2 for formal definitions). Unfortunately, as shown in [DDKT16, DKS16a], this “cover-based” approach cannot lead to qualitatively faster algorithms, due to a matching existential lower bound on the size of the corresponding ϵ -covers. In a related algorithmic work, Goldberg and Turchetta [GT15] studied two-strategy anonymous games ($k = 2$) and designed a polynomial-time algorithm that computes an ϵ -approximate Nash equilibria for $\epsilon = \Omega(n^{-1/4})$.

The aforementioned discussion prompts the following natural question: *What is the precise approximability of computing Nash equilibria in anonymous games?* In this paper, we make progress on this question by establishing the following result: For any $\delta > 0$, and any n -player anonymous game with a constant number of strategies, there exists a $\text{poly}_\delta(n)$ time algorithm that computes an ϵ -approximate Nash equilibrium of the game, for $\epsilon = 1/n^{1-\delta}$.² Moreover, we show that the existence of a polynomial-time algorithm that computes an ϵ -approximate Nash equilibrium for $\epsilon = 1/n^{1+\delta}$, for any small constant $\delta > 0$ – i.e., slightly better than the approximation guarantee of our algorithm – would imply the existence of a fully polynomial-time approximation scheme (FPTAS) for the problem. That is, we essentially show that the value $\epsilon = 1/n$ is the threshold for the polynomial-time approximability of Nash equilibria in anonymous games, unless there is an FPTAS. In the following subsection, we describe our results in detail and provide an overview of our techniques.

1.1 Our Results and Techniques We study the following question:

For n -player k -strategy anonymous games, how small can ϵ be (as a function of n), so that an ϵ -approximate Nash equilibrium can be computed in polynomial time?

Upper Bounds. We present two different algorithms (Theorems 1.1 and 1.2) for computing approximate Nash equilibria in anonymous games. Both algorithms run in polynomial time and compute ϵ -approximate equilibria for an inverse polynomial ϵ above a certain threshold.

THEOREM 1.1. (MAIN) *For any $\delta > 0$, and any n -player k -strategy anonymous game, there is a $\text{poly}_{\delta,k}(n)$ time algorithm that computes an $(1/n^{1-\delta})$ -approximate equilibrium of the game.*

²The runtime of our algorithm depends exponentially in $1/\delta$. We remind the reader that the algorithms of [DDKT16, DKS16a] run in quasi-polynomial time for any value of ϵ inverse polynomial in n .

THEOREM 1.2. *For any n -player k -strategy anonymous game, we can compute an $\tilde{O}(n^{-1/3}k^{11/3})$ -approximate equilibrium in time $\tilde{O}((n+k)kn^k)$.*

Prior to our work, for $k > 2$, no polynomial time ϵ -approximation was known for any inverse polynomial ϵ . For $k = 2$, the best previous result is due to [GT15] who gave a polynomial-time algorithm for $\epsilon = \Omega(n^{-1/4})$.

Overview of Techniques. The high-level idea of our approach is this: If the desired accuracy ϵ is above a certain threshold, we do not need to enumerate over an ϵ -cover for the set of all PMDs. Our approach is in part inspired by [GT15], who design an algorithm (for $k = 2$ and $\epsilon = \Omega(n^{-1/4})$) in which all players use one of the two pre-selected mixed strategies. We note that for $k = 2$, PMDs are tantamount to Poisson Binomial distributions (PBDs), i.e., sums of independent Bernoulli random variables. The [GT15] algorithm can be equivalently interpreted as guessing a PBD from an appropriately small set. One reason this idea succeeds is the following: If every player randomizes, then the variance of the resulting PBD must be relatively high, and (as a result) the corresponding subset of PBDs has a smaller cover.

Our quantitative improvement for the $k = 2$ case is obtained as follows: Instead of enforcing players to selected specific mixed strategies – as in [GT15] – we show that there always exists an ϵ -approximate equilibrium where the associated PBD has variance at least $\Theta(n\epsilon)$. When $\epsilon = n^{-c}$ for some $c < 1$, the variance is polynomial in n . We then construct a polynomial-size ϵ -cover for the subset of PBDs with variance at least this much, which leads to a polynomial-time algorithm for computing ϵ -approximate equilibria in 2-strategy anonymous games.

The idea for the general case of $k > 2$ is similar, but the details are more elaborate, since the structure of PMDs is more complicated for $k > 2$. We proceed as follows: We start by showing that there is an ϵ -approximate equilibrium whose corresponding PMD has a large variance in each direction. Our main structural result is a robust moment-matching lemma (Lemma 3.3), which states that the closeness in low-degree moments of two PMDs, with large variance in each direction, implies their closeness in total variation distance. The proof of this lemma uses Fourier analytic techniques, building on and strengthening previous work [DKS16a]. As a consequence of our moment-matching lemma, we can construct a polynomial-size $(\epsilon/5)$ -cover for PMDs with such large variance. We then iterate through this cover to find an ϵ -approximate equilibrium, using a dynamic programming approach similar to the one in [DP15].

We now provide a brief intuition of our moment-matching lemma. Intuitively, if the two PMDs in question are both very close to discrete Gaussians, then the closeness in the first *two* moments is sufficient. Lemma 3.3 can be viewed as a generalization of this intuition, which gives a quantitative tradeoff between the number of moments we need to approximately match and the size of the variance. The proof of Lemma 3.3 exploits the sparsity of the Fourier transform of our PMDs, and the fact that higher variance allows us to take fewer terms in the Taylor expansion when we use moments to approximate the logarithmic Fourier transform. This completes the proof sketch of Theorem 1.1.

Our second algorithm (Theorem 1.2) addresses the need to play simple strategies. Players tend to favor simple strategies which are easier to learn and implement, even if these strategies might have slightly sub-optimal payoffs [Sim82]. In addition, our algorithm is significantly faster in this case. We build on the idea of [GT15] to “smooth” an anonymous game by forcing all the players to randomize. We prove that the perturbed game is Lipschitz and therefore admits a pure Nash equilibrium, which corresponds to simple approximate equilibria of a specific form in the original game: Each player plays one strategy with probability $1 - \delta$ for some small δ , and plays other strategies uniformly at random with probability δ . To prove that the perturbed game is Lipschitz, we make essential use of the recently established multivariate central limit theorem (CLT) in Daskalakis *et al.* [DDKT16] and Diakonikolas, Kane and Stewart [DKS16a] to show that if we add more noise (corresponding to $\delta = \Theta(n^{-1/3})$), the associated PMD is sufficiently close to a discrete Gaussian.

Lower Bounds. When $\epsilon = 1/n$, we can show that there is an ϵ -approximate equilibrium where the associated PMD has a variance at least $1/k$ in every direction. Unfortunately, the PMDs in the explicit quasi-polynomial-size lower bounds given in [DDKT16, DKS16a] satisfy this property. Thus, we need a different approach to get a polynomial-time algorithm for $\epsilon = 1/n$ or smaller.

In fact, we prove the following results, which states that even a slight improvement of our upper bound in Theorem 1.1 would imply an FPTAS for computing Nash equilibria in anonymous games. It is important to note that Theorem 1.3 applies to all algorithms, not only the ones that leverage the structure of PMDs.

THEOREM 1.3. *For n -player k -strategy anonymous games with $k = O(1)$, if we can compute an $O(n^{-c})$ -approximate equilibrium in polynomial time for some constant $c > 1$, then there is an FPTAS³ for comput-*

ing (well-supported) Nash equilibria of k -strategy anonymous games.

Remark. As observed in [DDKT16], because there is a quasi-polynomial time algorithm for computing an (n^{-c}) -approximate equilibrium in anonymous games, the problem cannot be PPAD-Complete unless $\text{PPAD} \subseteq \text{Quasi-PTIME}$. On the other hand, we do not know how to improve the quasi-polynomial-time upper bounds of [DDKT16, DKS16a] when $\epsilon < 1/n$.

Recall that computing an ϵ -approximate equilibrium of a two-player general-sum $n \times n$ game (2-NASH) for constant ϵ also admits a quasi-polynomial-time algorithm [LMM03]. Very recently, Rubinstein [Rub16] showed that, assuming the exponential time hypothesis (ETH) for PPAD, for some sufficiently small universal constant $\epsilon > 0$, quasi-polynomial-time is necessary to compute an ϵ -approximate equilibrium of 2-NASH. It is a plausible conjecture that quasi-polynomial-time is also required for ϵ -Nash equilibria in anonymous games, when $\epsilon = n^{-c}$ for some constant $c > 1$. In particular, this would imply that there is no FPTAS for computing approximate Nash equilibria in anonymous games, and consequently the upper bound of Theorem 1.1 is essentially tight.

2 Notation and Background

Anonymous Games. We study anonymous games $(n, k, \{u_a^i\}_{i \in [n], a \in [k]})$ with n players labeled by $[n] = \{1, \dots, n\}$, and k common strategies labeled by $[k]$ for each player. The payoff of a player depends on her own strategy, and how many of her peers choose which strategy, but not on their identities. When player $i \in [n]$ plays strategy $a \in [k]$, her payoffs are given by a function u_a^i that maps the possible outcomes (partitions of all other players) Π_{n-1}^k to the interval $[0, 1]$, where $\Pi_{n-1}^k = \{(x_1, \dots, x_k) \mid x_j \in \mathbb{Z}_+ \wedge \sum_{j=1}^k x_j = n - 1\}$.

Approximate Equilibria. We denote by Δ_S a distribution on the set S . A *mixed strategy* is an element of $\Delta_{[k]}$, and a *mixed strategy profile* $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$ maps every player i to her mixed strategy $\mathbf{s}_i \in \Delta_{[k]}$. We use \mathbf{s}_{-i} to denote the strategies of players other than i in \mathbf{s} .

A mixed strategy profile \mathbf{s} is an ϵ -approximate Nash equilibrium for some $\epsilon \geq 0$ iff $\forall i \in [n], \forall a' \in [k]$,

$$\mathbb{E}_{x \sim \mathbf{s}_{-i}} [u_{a'}^i(x)] \leq \mathbb{E}_{x \sim \mathbf{s}_{-i}, a \sim \mathbf{s}_i} [u_a^i(x)] + \epsilon,$$

where $x \in \Pi_{n-1}^k$ is the partition formed by $n - 1$ random samples (independently) drawn from $[k]$ according

³A fully polynomial-time approximation scheme (FPTAS) is an algorithm that runs in time $\text{poly}(n, 1/\epsilon)$ and returns an ϵ -optimal solution, or in our context, returns an ϵ -approximate Nash equilibrium.

³A fully polynomial-time approximation scheme (FPTAS) is

to the distributions \mathbf{s}_{-i} . Note that given a mixed strategy profile \mathbf{s} , we can compute a player's expected payoff to precision ϵ in time $\text{poly}(n^k \log(1/\epsilon))$ by straightforward dynamic programming, and hence throughout this paper we assume that we can compute players' payoffs exactly given their mixed strategies.

Poisson Multinomial Distributions. A k -Categorical Random Variable (k -CRV) is a vector random variable supported on the set of k -dimensional basis vectors $\{e_1, \dots, e_k\}$. A k -CRV is i -maximal if e_i is its most likely outcome (break ties by taking the smallest index i). A k -Poisson Multinomial Distribution of order n , or an (n, k) -PMD, is a vector random variable of the form $X = \sum_{i=1}^n X_i$ where the X_i 's are independent k -CRVs. The case of $k = 2$ is usually referred to as Poisson Binomial Distribution (PBD).

Note that a mixed strategy profile $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$ of an n -player k -strategy anonymous game corresponds to the k -CRVs $\{X_1, \dots, X_n\}$ where $\Pr[X_i = e_a] = \mathbf{s}_i(a)$. The expected payoff of player $i \in [n]$ for playing pure strategy $a \in [k]$ can also be written as $\mathbb{E}[u_a^i(X_{-i})] = \mathbb{E}\left[u_a^i\left(\sum_{j \neq i, j \in [n]} X_j\right)\right]$.

Let $X = \sum_{i=1}^n X_i$ be an (n, k) -PMD such that for $i \in [n]$ and $j \in [k]$ we denote $p_{i,j} = \Pr[X_i = e_j]$, where $\sum_{j=1}^k p_{i,j} = 1$. For $m = (m_1, \dots, m_k) \in \mathbb{Z}_+^k$, we define the m^{th} -parameter moments of X to be $M_m(X) \stackrel{\text{def}}{=} \sum_{i=1}^n \prod_{j=1}^k p_{i,j}^{m_j}$. We refer to $\|m\|_1 = \sum_{j=1}^k m_j$ as the degree of the parameter moment $M_m(X)$.

Total Variation Distance and Covers.

The total variation distance between two distributions P and Q supported on a finite domain A is $d_{\text{TV}}(P, Q) \stackrel{\text{def}}{=} \max_{S \subseteq A} |P(S) - Q(S)| = (1/2)\|P - Q\|_1$.

If X and Y are two random variables ranging over a finite set, their total variation distance $d_{\text{TV}}(X, Y)$ is defined as the total variation distance between their distributions. For convenience, we will often blur the distinction between a random variable and its distribution.

Let (\mathcal{X}, d) be a metric space. Given $\epsilon > 0$, a subset $\mathcal{Y} \subseteq \mathcal{X}$ is said to be a proper ϵ -cover of \mathcal{X} with respect to the metric $d : \mathcal{X}^2 \rightarrow \mathbb{R}_+$, if for every $X \in \mathcal{X}$ there exists some $Y \in \mathcal{Y}$ such that $d(X, Y) \leq \epsilon$. In this work, we will be interested in constructing ϵ -covers for high-variance PMDs under the total variation distance metric.

Multidimensional Fourier Transform. For $x \in \mathbb{R}$, we will denote $e(x) \stackrel{\text{def}}{=} \exp(-2\pi i x)$. The (continuous) Fourier Transform of a function $F : \mathbb{Z} \rightarrow \mathbb{C}$ is the function $\hat{F} : [0, 1]^k \rightarrow \mathbb{C}$ defined as $\hat{F}(\xi) = \sum_{x \in \mathbb{Z}^k} e(\xi \cdot x) F(x)$. For the case that F is a probability mass function, we can equivalently write $\hat{F}(\xi) = \mathbb{E}_{x \sim F}[e(\xi \cdot x)]$.

Let $X = \sum_{i=1}^n X_i$ be an (n, k) -PMD with $p_{i,j} \stackrel{\text{def}}{=}$

$\Pr[X_i = e_j]$. To avoid clutter in the notation, we will use the symbol X to denote the corresponding probability mass function. With this convention, we can write that $\hat{X}(\xi) = \prod_{i=1}^n \hat{X}_i(\xi) = \prod_{i=1}^n \sum_{j=1}^k e(\xi_j) p_{i,j}$.

3 Searching Fewer Moments: Theorem 1.1

In this section, we present a polynomial-time algorithm that, for n -player anonymous games with bounded number of strategies, computes an ϵ -approximate equilibrium with $\epsilon = n^{-c}$ for any constant $c < 1$. As a warm up, we start by describing the simpler setting of two-strategy anonymous games ($k = 2$). The main results of this section is Theorem 1.1 that applies to general k -strategy anonymous games for any constant $k \geq 2$.

At a high level, we first prove the existence of ϵ -approximate Nash equilibria in which the corresponding PMDs have high variance and every player randomizes (Lemma 3.1). We then use our robust moment matching lemma (Lemma 3.3) to show that when two PMDs have high variances, the closeness in their constant-degree parameter moments implies their closeness in total variation distance. The fact that matching the constant-degree moments suffices allows us to construct a polynomial-size $(\epsilon/5)$ -cover for set subset of all PMDs with large variance. We then iterate through this cover to find an ϵ -approximate equilibrium (Algorithm 2).

LEMMA 3.1. *For an n -player k -strategy anonymous game, there always exists an ϵ -approximate equilibrium where every player plays each strategy with probability at least $\frac{\epsilon}{k-1}$.*

Proof. Given an anonymous game $G = (n, k, \{u_a^i\}_{i \in [n], a \in [k]})$, we smooth players' utility functions by requiring every player to randomize. Fix $\epsilon > 0$, we define an ϵ -perturbed game G_ϵ as follows. When a player plays some pure strategy $a \in [k]$ in G_ϵ , we map it back to the original game as if she plays strategy j with probability $1 - \epsilon$, and plays some other strategy $a' \neq a$ uniformly at random (i.e., she plays a' with probability $\frac{\epsilon}{k-1}$). Her payoff in G_ϵ also accounts for such perturbation, and is defined to be her expected payoff given that all the players (including herself) would deviate to other strategies uniformly at random with probability ϵ .

Formally, let $X_\epsilon(e_j)$ denote the k -CRV that takes value e_j with probability $1 - \epsilon$, and takes value $e_{j'}$ with probability $\frac{\epsilon}{k-1}$ for each $j' \neq j$. The payoff structure of G_ϵ is given by

$$u_a^i(x) = (1 - \epsilon)\mathbb{E}[u_a^i(M_\epsilon(x))] + \frac{\epsilon}{k-1} \sum_{a' \neq a} \mathbb{E}[u_{a'}^i(M_\epsilon(x))] \\ \forall i \in [n], a \in [k], x \in \Pi_{n-1}^k,$$

where $M_\epsilon(x) = \sum_{j \in [k]} x_j X_\epsilon(e_j)$ is an $(n-1, k)$ -PMD that corresponds to the perturbed outcome of the partition $x \in \Pi_{n-1}^k$ of all other players.

Let $\mathbf{s}' = (\mathbf{s}'_1, \dots, \mathbf{s}'_n)$ denote any exact Nash equilibrium of G_ϵ . We can interpret this mixed strategy profile in G equivalently as $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$, where $\mathbf{s}_i = (1 - \frac{k\epsilon}{k-1})\mathbf{s}'_i + \frac{\epsilon}{k-1}\mathbf{1}$, where $\mathbf{1} = (1, \dots, 1)$. We know that under \mathbf{s} each player has no incentive to deviate to the mixed strategies $X_\epsilon(e_j)$ for all $j \in [k]$, therefore a player can gain at most ϵ by deviating to pure strategies in G , so \mathbf{s} is an ϵ -approximate equilibrium with $s_i(j) \geq \frac{\epsilon}{k-1}$ for all $i \in [n], j \in [k]$.

Warm-up: The Case of $k = 2$ Strategies. For two-strategy anonymous games ($k = 2$), if all the players put at least ϵ probability mass on both strategies, the resulting PBD is going to have variance at least $n\epsilon(1-\epsilon)$. When $\epsilon = n^{-c}$ for some constant $c < 1$, the variance is at least $\Theta(n^{1-c}) = n^{\Theta(1)}$. We can now use the following lemma from [DKS16c], which states that if two PBDs P and Q are close in the first few moments, then P and Q are ϵ -close in total variation distance. Note that without any assumption on the variance of the PBDs, we would need to check the first $O(\log(1/\epsilon))$ moments, but when the variance is $n^{\Omega(1)}$, which is the case in our application, we only need the first constant number of moments to match.

LEMMA 3.2. ([DKS16c]) *Let $\epsilon > 0$. Let P and Q be n -PBDs with P having parameters $p_1, \dots, p_s \leq 1/2$ and $p'_1, \dots, p'_{s'} > 1/2$, and Q having parameters $q_1, \dots, q_s \leq 1/2$ and $q'_1, \dots, q'_{s'} > 1/2$. Suppose that $V = \text{Var}[P] + 1 = \Theta(\text{Var}[Q] + 1)$ and let $C > 0$ be a sufficiently large constant. Suppose furthermore that for $A = C\sqrt{\log(1/\epsilon)/V}$ and for all positive integers ℓ it holds*

$$(3.1) \quad A^\ell \left(\left| \sum_{i=1}^s p_i^\ell - \sum_{i=1}^s q_i^\ell \right| + \left| \sum_{i=1}^{s'} (1-p'_i)^\ell - \sum_{i=1}^{s'} (1-q'_i)^\ell \right| \right) < \frac{\epsilon}{C \log(1/\epsilon)}$$

Then $d_{\text{TV}}(P, Q) < \epsilon$.

Let $\epsilon = n^{-c}$. For Lemma 3.2 we have $V \geq n\epsilon(1-\epsilon)$ and $A = \Theta(\sqrt{\log(1/\epsilon)/V}) = O\left(\sqrt{\frac{\log n}{n^{1-c}}}\right)$. The difference in the moments of parameters of P and Q in Equation (3.1) is bounded from above by n , so whenever $\ell > \frac{2+2c}{1-c}$, the condition in Lemma 3.2 is automatically satisfied for sufficiently large n because

$$A^\ell n = O\left(\frac{\log^{\ell/2} n}{n^{(1-c)\ell/2}}\right) < \frac{1}{C \cdot n^c \cdot c \log n} = \frac{\epsilon}{C \log(1/\epsilon)}.$$

So it is enough to search over the first $\ell = \Theta\left(\frac{1}{1-c}\right)$ moments when each player put probability at least $\Omega(n^{-c})$ on both strategies. The algorithm for finding such an ϵ -approximate equilibrium uses moment search and dynamic programming, and is given for the case of general k in the remainder of this section.

The General Case: k Strategies. We now present our algorithm for n -player anonymous games with $k > 2$ strategies and prove Theorem 1.1. The intuition of the $k = 2$ case carries over to the general case, but the details are more elaborate. First, we show (Claim 3.1) that there exists an ϵ -approximate equilibrium whose corresponding PMD has variance $(n\epsilon/k)$ in all directions orthogonal to the vector $\mathbf{1} = (1, \dots, 1)$. Then, we prove (Lemma 3.3) that when two PMDs have such high variances, the closeness in their constant-degree parameter moments translates to their closeness in total variation distance. This structural result allows us to build a polynomial-size cover for all PMDs with high variance, which leads to a polynomial-time algorithm for computing ϵ -approximate Nash equilibria (Algorithm 2).

We first prove that when all the players put probability at least $\frac{\epsilon}{k-1}$ on each strategy, the covariance matrix of the resulting PMD has relatively large eigenvalues, except the zero eigenvalue associated with the all-one eigenvector. The all-one eigenvector has eigenvalue zero because the coordinates of X always sum to n .

CLAIM 3.1. *Let $X = \sum_{i=1}^n X_i$ be an (n, k) -PMD and let Σ be the covariance matrix of X . If $p_{i,j} = \Pr[X_i = e_j] \geq \frac{\epsilon}{k-1}$ for all $i \in [n]$ and $j \in [k]$, then all eigenvalues of Σ but one are at least $\frac{n\epsilon}{k-1}$.*

Proof. For any unit vector $v \in \mathbb{R}^k$ that is orthogonal to the all-one vector $\mathbf{1}$, i.e., $\sum_j v_j = 0$ and $\sum_j v_j^2 = 1$, combining this with the assumption that $p_{i,j} \geq \frac{\epsilon}{k-1}$ we have,

$$\begin{aligned} \text{Var}[v^T X_i] &= \mathbb{E}\left[\left(v^T X_i - \mathbb{E}[v^T X_i]\right)^2\right] \\ &= \sum_{j=1}^n p_{i,j} \left(v_j - \sum_{j'=1}^n p_{i,j'} v_{j'}\right)^2 \\ &\geq \min_j \{p_{i,j}\} \cdot \left(1 + n \left(\sum_{j'=1}^n p_{i,j'} v_{j'}\right)^2\right) \\ &\geq \frac{\epsilon}{k-1}. \end{aligned}$$

Therefore,

$$v^T \Sigma v = \text{Var}[v^T X] = \sum_{i=1}^n \text{Var}[v^T X_i] \geq \frac{n\epsilon}{k-1}.$$

So, for all eigenvectors v orthogonal to $\mathbf{1}$, we have $v^T \Sigma v = \lambda v^T v = \lambda \geq \frac{n\epsilon}{k-1}$ as claimed.

The following robust moment-matching lemma provides a bound on how close degree ℓ moments need to be so that two (n, k) -PMDs are ϵ -close to each other, under the assumption that $n \gg k$ (the anonymous game has many players and few strategies) and $p_{i,j} \geq \frac{\epsilon}{k-1}$ (every player randomizes). Lemma 3.3 allows us to build a polynomial-size $(\epsilon/5)$ -cover for PMDs with high variance, and since we know that there is an ϵ -approximate equilibrium with a high variance, we are guaranteed to find one in our cover.

LEMMA 3.3. *Fix $0 < c < 1$ and let $\epsilon = n^{-c}$. Assume that $n \geq k^{\Theta(k)}$ for some sufficiently large constant in the exponent. Let X, Y be (n, k) -PMDs with $X = \sum_{i=1}^k X^i$, $Y = \sum_{i=1}^k Y^i$ where each X^i, Y^i is an i -maximal PMD. Let Σ and Σ' denote the covariance matrices of X and Y respectively. If all eigenvalues of Σ, Σ' but one are at least $\epsilon n/k$, and for $\ell \leq \frac{2+2c}{1-c}$ all the parameter moments m of degree ℓ satisfy that*

$$|M_m(X^i) - M_m(Y^i)| \leq n^{-c}.$$

Then, we have that $d_{TV}(X, Y) \leq \epsilon$.

Lemma 3.3 follows from the next proposition whose proof is given in the following subsection.

PROPOSITION 3.1. *Let $\epsilon > 0$. Let X, Y be (n, k) -PMDs with $X = \sum_{i=1}^k X^i$, $Y = \sum_{i=1}^k Y^i$ where each X^i, Y^i is an i -maximal PMD. Let Σ and Σ' denote the covariance matrices of X and Y respectively, where all eigenvalues of Σ and Σ' but one are at least σ^2 , where $\sigma \geq \text{poly}(k \log(1/\epsilon))$. Suppose that for $1 \leq i \leq k$, $\ell \geq 1$, and for all moments m of degree ℓ with $m_i = 0$, the following holds for a sufficiently large constant C' ,*

$$|M_m(X^i) - M_m(Y^i)| \leq \frac{\epsilon \cdot \sigma^\ell}{C'^{k+\ell} \cdot k^{3\ell/2+1} \cdot \log^{k+\ell/2}(1/\epsilon)}$$

Then $d_{TV}(X, Y) \leq \epsilon$.

The proof of Proposition 3.1 exploits the sparsity of the continuous Fourier transform of our PMDs, as well as careful Taylor approximations of the logarithm of the Fourier transform.

Proof. [Proof of Lemma 3.3 from Proposition 3.1] In order to guarantee that $d_{TV}(X, Y) \leq \epsilon$, Proposition 3.1 requires the following condition to hold for a sufficiently large constant C' :

$$(3.2) \quad \forall i \in [k], \ell \geq 1, \quad |M_m(X^i) - M_m(Y^i)| \leq \frac{\epsilon}{k(C' \log(1/\epsilon))^k} \cdot \left(\frac{\sqrt{\epsilon n/k}}{C' k^{3/2} \log^{1/2}(1/\epsilon)} \right)^\ell.$$

To prove the lemma, we use the fact that $n \gg k$ and essentially ignore all the terms except polynomials of n . Formally, we first need to show that

$$\frac{\epsilon}{k(C' \log(1/\epsilon))^k} \cdot \left(\frac{\sqrt{\epsilon n/k}}{C' k^{3/2} \log^{1/2}(1/\epsilon)} \right)^\ell \geq n^{-c}, \quad \forall \ell \geq 1,$$

under the assumption that $c < 1$, $\epsilon = n^{-c}$ and $n \geq k^{O(k/(1-c))}$. After substituting $\epsilon = n^{-c}$, observe that $n^{1-c} \geq C'^2 k^4 \log n$, so the term inside the ℓ -th power is greater than 1. Thus, we only need to check this inequality for $\ell = 1$, which simplifies to $n^{1-c} \geq C'^2 k^4 \log n$ and holds true.

In addition, we need to show that condition (3.2) holds automatically for $\ell > \frac{2+2c}{1-c}$. This follows from the fact that the difference in parameter moments is at most n and $n \gg k$,

$$\begin{aligned} \forall \ell > \frac{2+2c}{1-c}, \quad |M_m(X^i) - M_m(Y^i)| &\leq n \\ &\leq \frac{\epsilon}{k(C' \log(1/\epsilon))^k} \cdot \left(\frac{\sqrt{\epsilon n/k}}{C' k^{3/2} \log^{1/2}(1/\epsilon)} \right)^\ell. \end{aligned}$$

We recall some of the notations for readability before we describe the construction of our ϵ -cover of high-variance PMDs. We use X to denote a generic (ℓ, k) -PMD for some $\ell \in [n]$, and we denote $p_{i,j} = \Pr[X_i = e_j]$. We use $A_t \subseteq [\ell]$ to denote the set of t -maximal CRVs in X , where a k -CRV is t -maximal if e_t is its most likely outcome, and we use $X^t = \sum_{i \in A_t} X_i$ to denote the t -maximal component PMD of X . For a vector $m = (m_1, \dots, m_k) \in \mathbb{Z}_+^k$, we define m^{th} parameter moment of X^t to be $M_m(X^t) = \sum_{i \in A_t} \prod_{j=1}^k p_{i,j}^{m_j}$. We refer to $\|m\|_1 = \sum_{j=1}^k m_j$ as the degree of $M_m(X)$. We use \mathcal{S} to denote the set of all k -CRVs whose probabilities are multiples of $\frac{\epsilon}{20kn}$.

Lemma 3.3 states that the high-degree parameter moments match automatically, which allows us to impose an appropriate grid on the low-degree moments to cover the set of high-variance PMDs. The size of this cover can be bounded by a simple counting argument: We have at most $k^{O(\frac{1}{1-c})}$ moments with degree at most $O(\frac{1}{1-c})$, and we need to approximate these moments for each t -maximal component PMDs, so there are at most $k \cdot k^{O(\frac{1}{1-c})} = k^{O(\frac{1}{1-c})}$ moments $M_m(X^t)$ that we care about. We approximate these moments to precision n^{-c} , and the moments are at most n , so the size of the cover is $\left(\frac{n}{n^{-c}}\right)^{k^{O(\frac{1}{1-c})}} = n^{k^{O(1/(1-c))}}$.

We define this grid on low-degree moments formally in the following lemma. For every (ℓ, k) -PMD X with $\ell \in [n]$, we associate some data $D(X)$ with X , which is a vector of the approximate values of the low-degree moments $M_m(X^t)$ of X .

Algorithm 1: GenerateData

Input : $\{\mathcal{S}_i\}_{i=1}^n, \epsilon > 0$.
Output: The set of all possible data \mathcal{D} of (n, k) -PMDs $X = \sum_{i=1}^n X_i$ where $X_i \in \mathcal{S}_i$.
 $\mathcal{D}_0 = \{0\}$;
for $\ell = 1 \dots n$ **do**
 forall the $D \in \mathcal{D}_{\ell-1}$ **do**
 forall the $W \in \mathcal{S}_\ell$ **do**
 Add $D + D(W)$ to \mathcal{D}_ℓ if it is not in \mathcal{D}_ℓ already;
 Keep track of an (ℓ, k) -PMD whose data is $D + D(W)$;
 end
 end
end
return $\mathcal{D} = \mathcal{D}_n$;

LEMMA 3.4. Fix $0 < c < 1$ and let $\epsilon = n^{-c}$. Assume that $n \geq k^{\Theta(k)}$ for some sufficiently large constant in the exponent. The data $D(W)$ of a k -CRV W as follows:

$$D(W)_{m,t} = \begin{cases} M_m(W) \text{ rounded to} & \text{if } W \text{ is} \\ \text{the nearest integer} & t\text{-maximal.} \\ \text{multiple of } n^{-c}/n, & \\ 0, & \text{otherwise.} \end{cases}$$

For $\ell \in [n]$, we define the data of an (ℓ, k) -PMD $X = \sum_{i=1}^\ell X_i$ to be the sum of the data of its k -CRVs: $D(X) = \sum_{i=1}^\ell D(X_i)$. The data $D(X)$ satisfies two important properties:

1. (Representative) If $D(X) = D(Y)$ for two (n, k) -PMDs (or two $(n-1, k)$ -PMDs) X and Y , then $d_{TV}(X, Y) \leq \epsilon$.
2. (Extensible) For independent PMDs X and Y , we have that $D(X + Y) = D(X) + D(Y)$.

Proof. The extensible property follows directly from the definition of $D(X)$. To see the representative property, note that we round $M_m(W)$ to the nearest integer multiple of n^{-c}/n , so the error in the moments of W is at most $n^{-c}/(2n)$. When we add up the data of an (n, k) -PMD or $(n-1, k)$ -PMD, the error in the moments of each t -maximal component PMDs is at most $n^{-c}/2$. So if two PMDs X and Y have the same data, their low-degree moments differ by at most n^{-c} , and then by Lemma 3.3 we have $d_{TV}(X, Y) \leq \epsilon$.

Our algorithm (Algorithm 2) for computing approximate equilibria is similar to the approach used

Algorithm 2: Moment Search

Input : An n -player k -strategy anonymous game G , $\epsilon = n^{-c}$ for some $c < 1$.
Output: An ϵ -approximate equilibrium of G .
 $\mathcal{D}_n = \text{GenerateData}(\{\mathcal{S}_i = \mathcal{S}\}_{i=1}^n, \epsilon/5)$;
 $\mathcal{D}_{n-1} = \text{GenerateData}(\{\mathcal{S}_i = \mathcal{S}\}_{i=1}^{n-1}, \epsilon/5)$;
forall the $D \in \mathcal{D}_n$ **do**
 Set $\mathcal{S}_i = \emptyset$ for all i ;
 forall the $X_i \in \mathcal{S}$ **do**
 Let $D_{-i} = D - D(X_i)$;
 if $\exists Y_{D_{-i}} \in \mathcal{D}_{n-1}$ with $D(Y_{D_{-i}}) = D_{-i}$ and X_i is a $(3\epsilon/5)$ -best response to $Y_{D_{-i}}$ **then**
 Add X_i to \mathcal{S}_i ;
 end
 end
 $\mathcal{D}'_n = \text{GenerateData}(\{\mathcal{S}_i\}_{i=1}^n, \epsilon/5)$;
 if $D \in \mathcal{D}'_n$ **then**
 return (X_1, \dots, X_n) in \mathcal{D}'_n with $D(\sum_{i=1}^n X_i) = D$;
 end
end

in [DP15] and [DKS16a]. We start by constructing a polynomial-sized $(\epsilon/5)$ -cover of high-variance PMDs (Algorithm 1), and then iterate over this cover. For each element in the cover, we compute the set of $(3\epsilon/5)$ -best-responses for each player, and then run the cover construction algorithm again, but this time we only allow each player to choose from her $(3\epsilon/5)$ -best-responses. If we could reconstruct a PMD whose moments are close enough to the one we started with, then we have found an ϵ -approximate Nash equilibrium.

Recall that a mixed strategy profile for a k -strategy anonymous game can be represented as a list of k -CRVs (X_1, \dots, X_n) , where X_i describes the mixed strategy of player i . Recall that (X_1, \dots, X_n) is an ϵ -approximate Nash equilibrium if for each player i we have $\mathbb{E}[u_{X_i}^i(X_{-i})] \geq \mathbb{E}[u_a^i(X_{-i})] - \epsilon$ for all $a \in [k]$, where $X_{-i} = \sum_{j \neq i} X_j$ is the distribution of the sum of other players strategies.

LEMMA 3.5. Fix an anonymous game $G = (n, k, \{u_a^i\}_{i \in [n], a \in [k]})$ with payoffs normalized to $[0, 1]$. Let (X_1, \dots, X_n) and (Y_1, \dots, Y_n) be two lists of k -CRVs. If X_i is a δ -best response to X_{-i} , and $d_{TV}(X_{-i}, Y_{-i}) \leq \epsilon$, then X_i is a $(\delta + 2\epsilon)$ -best response to Y_{-i} . Moreover, if (X_1, \dots, X_n) is a δ -approximate equilibrium, and $d_{TV}(X_i, Y_i) + d_{TV}(X_{-i}, Y_{-i}) \leq \epsilon$ for all $i \in [n]$, then (Y_1, \dots, Y_n) is a $(\delta + 2\epsilon)$ -approximate equilibrium.

Proof. Since $u_a^i(x) \in [0, 1]$ for all $a \in [k]$ and $x \in \Pi_{n-1}^k$, we have that $\forall i \in [n], a \in [k]$,

$$|\mathbb{E}[u_a^i(X_{-i})] - \mathbb{E}[u_a^i(Y_{-i})]| \leq d_{\text{TV}}(X_{-i}, Y_{-i}).$$

Therefore, if $d_{\text{TV}}(X_{-i}, Y_{-i}) \leq \epsilon$, and player i cannot deviate and gain more than δ when other players play X_{-i} , then she cannot gain more than $(\delta + 2\epsilon)$ when other players play Y_{-i} instead of X_{-i} . The second claim combines the inequality above with the fact that, if player i plays Y_i instead of X_i and the mixed strategies of other players remain the same, her payoff changes by at most $d_{\text{TV}}(X_i, Y_i)$. Formally,

$$|\mathbb{E}[u_{X_i}^i(Z_{-i})] - \mathbb{E}[u_{Y_i}^i(Z_{-i})]| \leq d_{\text{TV}}(X_i, Y_i), \\ \forall k\text{-CRV } X_i, Y_i, \forall (n-1, k)\text{-PMD } Z_{-i}.$$

The next lemma states that by rounding an $(\epsilon/10)$ -approximate equilibrium, we can obtain an $(\epsilon/5)$ -approximate equilibrium where all the probabilities are integer multiples of $\frac{\epsilon}{20kn}$.

CLAIM 3.2. *There is an $(\epsilon/5)$ -approximate Nash equilibrium (X_1, \dots, X_n) , such that for all $i \in [n]$ and $j \in [k]$, the probabilities $p_{i,j} = \Pr[X_i = e_j]$ are multiples of $\frac{\epsilon}{20kn}$, and also $p_{i,j} \geq \frac{\epsilon}{10k}$.*

Proof. We start with an $(\epsilon/10)$ -approximate Nash equilibrium (Y_1, \dots, Y_n) from Lemma 3.1 with $p_{i,j} \geq \frac{\epsilon}{10k}$, and then round the probabilities to integer multiples of $\frac{\epsilon}{10kn}$. We construct X_i from Y_i as follows: for every $j < k$, we set $\Pr[X_i = e_j]$ to be $\Pr[Y_i = e_j]$ rounded down to a multiple of $\frac{\epsilon}{20kn}$ and we set $\Pr[X_i = e_k] = 1 - \sum_{j < k} \Pr[X_i = e_j]$ so the probabilities sum to 1. By triangle inequality of total variation distance, for every i we have $d_{\text{TV}}(X_i, Y_i) \leq \frac{\epsilon}{20n}$ and $d_{\text{TV}}(X_{-i}, Y_{-i}) \leq \frac{\epsilon(n-1)}{20n}$. An application of Lemma 3.5 shows that (X_1, \dots, X_n) is an $(\epsilon/5)$ -approximate equilibrium.

We are now ready to prove Theorem 1.1.

Proof. [Proof of Theorem 1.1] We show that for any n -player k -strategy anonymous game, if both $c > 0$ and k are constants, then there is a $\text{poly}(n)$ time algorithm that computes an ϵ -approximate equilibrium for $\epsilon = 1/n^{1-c}$. If $n = k^{O(k)} = O(1)$, we use the algorithm in [DP07] which runs in time $n^{(1/\epsilon)^{\Omega(k)}} = O(1)$. So for the rest of the proof, we assume that $n \geq k^{\Theta(k)}$ as required in Lemma 3.3 and 3.4, and prove that Algorithm 2 always outputs an ϵ -approximate Nash equilibrium, and bound the running time.

We first show that the output (X_1, \dots, X_n) is an ϵ -approximate equilibrium. Recall that \mathcal{S} is the set of

all k -CRVs whose probabilities are multiples of $\frac{\epsilon}{20kn}$, and $\mathcal{S}_i \subseteq \mathcal{S}$ is the set of approximate best-responses of player i . When we put X_i in \mathcal{S}_i , we checked that X_i is a $(3\epsilon/5)$ -best response to $Y_{D_{-i}}$, note that $D(Y_{D_{-i}}) = D - D(X_i) = D(X_{-i})$, so by Lemma 3.4 $d_{\text{TV}}(X_{-i}, Y_{D_{-i}}) \leq \epsilon/5$ for all i . By Lemma 3.5, X_i is indeed an ϵ -best response to X_{-i} for all i .

Next we show the algorithm must always output something. By Claim 3.2 there exists an $(\epsilon/5)$ -approximate equilibrium X'_i with each $X'_i \in \mathcal{S}$. If the algorithm does not terminate successfully first, it eventually considers $D(X')$. Because X'_{-i} is an $(n-1, k)$ -PMD, the algorithm can find some $Y_{D_{-i}}$ with $D(Y_{D_{-i}}) = D(X') - D(X'_i) = D(X'_{-i})$, and by Lemma 3.4 we have $d_{\text{TV}}(X'_{-i}, Y_{D_{-i}}) \leq \epsilon/5$ for all i . Since X'_i is an $(\epsilon/5)$ -best response to X'_{-i} , Lemma 3.5 yields that X'_i is a $(3\epsilon/5)$ -best response to $Y_{D_{-i}}$, so we would add each X'_i to \mathcal{S}_i . Then our cover construction algorithm is guaranteed to generate a set of data that includes $D(X')$, and Algorithm 2 would produce an output.

Finally, we bound the running time of Algorithm 2. Let $N = O\left(n^{k^{O(1/(1-c))}}\right)$ denote the size of the $(\epsilon/5)$ -cover for the high-variance PMDs. The cover can be constructed in time $O(n \cdot N \cdot |S|)$ as we try to add one k -CRV from S in each step. We iterate through the cover, and for each element in the cover, we need to find the subset $\mathcal{S}_i \subseteq \mathcal{S}$ of $(3\epsilon/5)$ -best responses for player i , and then run the cover construction algorithm again using only the best responses $\{\mathcal{S}_i\}_{i=1}^n$. So the overall running time of the algorithm is $O(nN|S|) \cdot (\text{poly}(n^k)|S| + O(nN|S|)) = n^{k^{O(1/(1-c))}}$. When both $0 < c < 1$ and k are constants, the running time is polynomial in n .

3.1 Proof of Proposition 3.1 This subsection is devoted to the proof of Proposition 3.1. For two (n, k) -PMDs with variance at least σ^2 in each direction, Proposition 3.1 gives a quantitative bound on how close degree ℓ moments need to be (as a function of ϵ, σ, k and ℓ , but independent of n), in order for the two PMDs to be ϵ -close in total variation distance.

The proof of Proposition 3.1 exploits the sparsity of the continuous Fourier transforms of our PMDs, as well as careful Taylor approximations of the logarithm of the Fourier transform. The fact that our PMDs have large variance enables us to take fewer low-degree terms in the Taylor approximation. For technical reasons, we split our PMD as the sum of k independent component PMDs, $X = \sum_{i=1}^k X^i$, where all the k -CRVs in the component PMD X^i is i -maximal. Because the Fourier transform of X is the product of the Fourier transform of X^i , we can just bound the pointwise difference

between the logarithm of Fourier transform of each component PMD. One technicality is that since we have no assumption on the variances of the component PMDs X^i , their Fourier transforms may not be sparse, so it is crucial that we bound this difference only on the effective support of the Fourier transform of the entire PMD.

We start by considering a set S that includes the effective support of X (and Y when we show that the means are close):

LEMMA 3.6. *Let X be an (n, k) -PMD with mean μ and covariance matrix Σ , such that all the non-zero eigenvalues of Σ is at least σ^2 where $\sigma \geq \text{poly}(1/\epsilon)$. Let S be the set of points $x \in \mathbb{Z}^k$ where $(x - \mu)^T \mathbf{1} = 0$ and*

$$(x - \mu)^T (\Sigma + I)^{-1} (x - \mu) \leq (Ck \log(1/\epsilon)),$$

for some sufficiently large constant C . Then, $X \in S$ with probability at least $1 - \epsilon/2$, and

$$|S| = \sqrt{\det(\Sigma + I)} \cdot O(\log(1/\epsilon))^{k/2}.$$

Proof. Applying Lemma 5.2 of [DKS16a], we have that $(X - \mu)^T (\Sigma + I)^{-1} (X - \mu) = O(k \log(k/\epsilon))$ with probability at least $1 - \epsilon$. The set of integer coordinate points in this ellipsoid is the set S . Note that $|S|$ is equal to the volume of $S' = \{y \in \mathbb{R}^k : \exists x \in S \text{ with } \|y - x\|_\infty \leq 1/2\}$, because S' is the disjoint union of cubes of volume 1, one for each integer point. But S' is again contained in an ellipsoid with $(y - \mu)^T (\Sigma + I)^{-1} (y - \mu) = O(k \log(k/\epsilon))$, so $|S| = \text{Vol}(S') = \sqrt{\det(\Sigma + I)} \cdot O(\log(1/\epsilon))^{k/2}$.

Next we show that \hat{X} , the Fourier transform of X , has a relatively small effective support. We fold the effective support onto $[0, 1]^k$ to obtain the set T . We use $[x]$ to denote the additive distance of $x \in \mathbb{R}$ to the closest integer, i.e., $[x] = \min_{x' \in \mathbb{Z}} |x - x'|$.

LEMMA 3.7. *Let X be an (n, k) -PMD with mean μ and covariance matrix Σ , such that all the non-zero eigenvalues of Σ are at least σ^2 where $\sigma \geq \text{poly}(k \log(1/\epsilon))$. Let S be as above. Let \hat{X} be the Fourier transform of X . Let $T \stackrel{\text{def}}{=} \{\xi \in [0, 1]^k : \exists \xi' \in \xi + \mathbb{Z}^k \text{ with } \xi'^T \Sigma \xi' \leq Ck \log(1/\epsilon)\}$, for some sufficiently large constant C . Then, we have that*

$$(i) \text{ For } \xi \in T, \text{ and for all } 1 \leq i, j \leq k, [\xi_i - \xi_j] \leq 2\sqrt{Ck \log(1/\epsilon)}/\sigma.$$

$$(ii) \text{ Vol}(T)|S| = O(C \log(1/\epsilon))^k.$$

$$(iii) \int_{[0, 1]^k \setminus T} |\hat{X}(\xi)| d\xi \leq \epsilon/(2|S|).$$

Lemma 3.7 is a technical generalization of Lemma 5.5 of [DKS16a]. Its proof is deferred to Appendix A. This lemma establishes that the contribution to the Fourier transform \hat{X} coming from points outside of T is negligibly small. We then use the sparsity of the Fourier transform to show that, if two PMDs have Fourier transforms that are pointwise sufficiently close within the effective support T , then the two PMDs are close in total variation distance.

LEMMA 3.8. *Let X, Y, S, T be as above. If $|\hat{X}(\xi) - \hat{Y}(\xi)| \leq \epsilon(C' \log(1/\epsilon))^{-k}$ for all $\xi \in T$ and a sufficiently large constant C' , then $d_{\text{TV}}(X, Y) \leq \epsilon$.*

Proof. For any $x \in \mathbb{Z}^k$, taking the inverse Fourier transform, we have that $\Pr[X = x] = \int_{\xi \in [0, 1]^k} e(-\xi \cdot x) \hat{X}(\xi) d\xi$ and similarly $\Pr[Y = x] = \int_{\xi \in [0, 1]^k} e(-\xi \cdot x) \hat{Y}(\xi) d\xi$. Thus,

$$\begin{aligned} & |\Pr[X = x] - \Pr[Y = x]| \\ &= \left| \int_{\xi \in [0, 1]^k} e(-\xi \cdot x) (\hat{X}(\xi) - \hat{Y}(\xi)) d\xi \right| \\ &\leq \int_{\xi \in [0, 1]^k} |\hat{X}(\xi) - \hat{Y}(\xi)| d\xi \\ &= \int_{\xi \in T} |\hat{X}(\xi) - \hat{Y}(\xi)| d\xi + \int_{\xi \in [0, 1]^k \setminus T} |\hat{X}(\xi) - \hat{Y}(\xi)| d\xi \\ &\leq \text{Vol}(T) \cdot \epsilon(C' \log(1/\epsilon))^{-k} + \frac{\epsilon}{2|S|} \\ &\leq \frac{O(C \log(1/\epsilon))^k}{|S|} \cdot \epsilon(C' \log(1/\epsilon))^{-k} + \frac{\epsilon}{2|S|} \\ &\leq \frac{\epsilon}{|S|}. \end{aligned}$$

Since X and Y are outside of S each with probability less than $\epsilon/2$, we have that $d_{\text{TV}}(X, Y) \leq \epsilon/2 + \frac{1}{2} \sum_{x \in S} |\Pr[X = x] - \Pr[Y = x]| \leq \epsilon$.

We now have all the ingredients to prove Proposition 3.1. For two PMDs X and Y that are close in their low-degree moments, we show that their Fourier transforms \hat{X} and \hat{Y} are pointwise close on T , and then by Lemma 3.8, X and Y are close in total variation distance.

Proof. [Proof of Proposition 3.1] Let X, Y, S, T be as above. Given Lemma 3.8, we only need to show that $\forall \xi \in T, |\hat{X}(\xi) - \hat{Y}(\xi)| \leq \epsilon(C' \log(1/\epsilon))^{-k}$.

Fix $\xi \in T$. We first examine, without loss of generality, the Fourier transform \hat{X}^k of the k -maximal

component PMD. Let $A_k \subseteq [n]$ denote the set of k -maximal CRVs.

$$\begin{aligned}
(3.3) \quad \widehat{X}^k(\xi) &= \prod_{i \in A_k} \sum_{j=1}^k e(\xi_j) p_{i,j} \\
&= e(|A_k| \xi_k) \prod_{i \in A_k} \left(1 - \sum_{j=1}^{k-1} (1 - e(\xi_j - \xi_k)) p_{i,j} \right) \\
&= e(|A_k| \xi_k) \exp \left(\sum_{i \in A_k} \log \left(1 - \sum_{j=1}^{k-1} (1 - e(\xi_j - \xi_k)) p_{i,j} \right) \right) \\
&= e(|A_k| \xi_k) \exp \left(- \sum_{i \in A_k} \sum_{\ell=1}^{\infty} \frac{1}{\ell} \left(\sum_{j=1}^{k-1} (1 - e(\xi_j - \xi_k)) p_{i,j} \right)^\ell \right) \\
&= e(|A_k| \xi_k) \cdot \exp \left(- \sum_{m \in \mathbb{Z}_+^{k-1}} \binom{\|m\|_1}{m} \frac{1}{\|m\|_1} M_m(X^k) \prod_{j=1}^{k-1} (1 - e(\xi_j - \xi_k))^{m_j} \right)
\end{aligned}$$

For notational convenience, we use Ψ_X^k to denote the expression inside $\exp(\cdot)$ on the last line of Equation (3.3). A similar formula holds for the Fourier transform \widehat{X}^i and \widehat{Y}^i of other i -maximal PMDs, and we use Ψ_X^i and Ψ_Y^i to denote the corresponding expressions inside $\exp(\cdot)$. Since the Fourier transform of a PMD is the product of the Fourier transform of its component PMDs, we have

$$\begin{aligned}
|\widehat{X}(\xi) - \widehat{Y}(\xi)| &= \left| \prod_{t=1}^k \widehat{X}^t(\xi) - \prod_{t=1}^k \widehat{Y}^t(\xi) \right| \\
&= \left| e \left(\sum_{t=1}^k |A_t| \xi_t \right) \prod_{t=1}^k (\exp(\Psi_X^t) - \exp(\Psi_Y^t)) \right| \\
&\leq 2\pi \sum_{t=1}^k |\Psi_X^t - \Psi_Y^t|,
\end{aligned}$$

where the last inequality is due to $e(\sum_{t=1}^k |A_t| \xi_t) = 1$, and $|\exp(a) - \exp(b)| \leq |a - b|$ if the real parts of a and b satisfy $\text{Re}(a), \text{Re}(b) \leq 0$.

So to prove that $\widehat{X}(\xi)$ and $\widehat{Y}(\xi)$ are pointwise close for all $\xi \in T$, it is enough to bound from above $2\pi \sum_{t=1}^k |\Psi_X^t - \Psi_Y^t|$. We use the fact that $|1 - e(\xi_j - \xi_k)| = O(|\xi_j - \xi_k|)$, and recall that $[\xi_i - \xi_j] \leq 2\sqrt{Ck \log(1/\epsilon)}/\sigma$ by Lemma 3.7. We also use the multinomial identity $\sum_{m \in \mathbb{Z}_+^{k-1}, \|m\|_1 = \ell} \binom{\ell}{m} = (k-1)^\ell$. When

C' is a sufficiently large constant, we have

$$\begin{aligned}
&|\widehat{X}(\xi) - \widehat{Y}(\xi)| \\
&\leq 2\pi \sum_{t=1}^k |\Psi_X^t - \Psi_Y^t| \\
&= 2\pi \sum_{t=1}^k \sum_{m \in \mathbb{Z}_+^{k-1}} \left(\binom{\|m\|_1}{m} \frac{1}{\|m\|_1} \right. \\
&\quad \left. |M_m(X^t) - M_m(Y^t)| \prod_{j=1}^{k-1} (1 - e(\xi_j - \xi_k))^{m_j} \right) \\
&\leq 2\pi \sum_{\ell=1}^{\infty} \left(\frac{(k-1)^\ell}{\ell} \left(O\left(\frac{\sqrt{k \log(1/\epsilon)}}{\sigma} \right) \right)^\ell \right. \\
&\quad \left. \sum_{t=1}^k \max_{m \in \mathbb{Z}_+^{k-1}, \|m\|_1 = \ell} |M_m(X^t) - M_m(Y^t)| \right) \\
&\leq \sum_{\ell=1}^{\infty} \left(k^\ell \left(\frac{C' \sqrt{k \log(1/\epsilon)}}{2\sigma} \right)^\ell k \right. \\
&\quad \left. \frac{\epsilon \sigma^\ell}{C'^{k+\ell} \cdot k^{3\ell/2+1} \cdot \log^{k+\ell/2}(1/\epsilon)} \right) \\
&= \sum_{\ell=1}^{\infty} 2^{-\ell} \epsilon (C' \log(1/\epsilon))^{-k} \\
&= \epsilon (C' \log(1/\epsilon))^{-k}.
\end{aligned}$$

4 Reductions: Proof of Theorem 1.3

In this section, we show that even a slight improvement of our upper bound would imply an FPTAS for computing (well-supported) Nash equilibria in anonymous games (Theorem 1.3). It is a plausible conjecture that assuming ETH for PPAD, there is no such FPTAS, in which case our upper bound (Theorem 1.1) is essentially tight.

Theorem 1.3 follows directly from the following two lemmas. Lemma 4.1 converts an $\frac{\epsilon^2}{4n}$ -approximate Nash equilibrium into an ϵ -well-supported Nash equilibrium⁴, by reallocating each player's probabilities on strategies with low expected payoffs to the best-response strategy (first observed in [DGP09]). Lemma 4.2 then uses a padding argument to show that, for ϵ -well-supported Nash equilibrium, the question of whether there is a

⁴A mixed strategy profile s is a *well-supported* Nash equilibrium iff $\forall i \in [n], \forall a, a' \in [k]$, we have $\mathbb{E}_{x \sim s_{-i}}[u_a^i(x)] > \mathbb{E}_{x \sim s_{-i}}[u_{a'}^i(x)] + \epsilon \implies s_i(a') = 0$, i.e., players can only put non-zero probability on ϵ -best-response strategies.

polynomial-time algorithm for $\epsilon = n^{-c}$ is equivalent for all constants $c > 0$.

LEMMA 4.1. *For any n -player game whose payoffs are normalized to be between $[0, 1]$, if we have an oracle for computing players' payoffs, we can efficiently convert an $\frac{\epsilon^2}{4n}$ -approximate equilibrium into an ϵ -well-supported equilibrium.*

Proof. Take an $\frac{\epsilon^2}{4n}$ -approximate equilibrium of the game. We call a strategy “good” for a player if the strategy is an $\frac{\epsilon}{2}$ -best response for the player, and we call it “bad” otherwise. A player can put at most probability $\frac{\epsilon}{2n}$ on the “bad” strategies without violating the $\frac{\epsilon^2}{4n}$ -approximate equilibrium condition. We move all the probabilities on “bad” strategies for all players to (any one of) their best responses simultaneously. After moving the probabilities, every player assigns non-zero probabilities only to the “good” strategies. Since the total probability we moved is at most $\frac{\epsilon}{2}$ and the payoffs are in $[0, 1]$, the previously “good” strategies ($\frac{\epsilon}{2}$ -best responses) are now ϵ -best responses.

LEMMA 4.2. *For n -player k -strategy anonymous games with $k = O(1)$, if an $\frac{1}{n^\gamma}$ -well-supported equilibrium can be computed in time $O(n^d)$ for constants $\gamma, d > 0$, then there is an FPTAS for computing approximate-well-supported Nash equilibria in anonymous games.*

Proof. Let ϵ be the desired quality of the well-supported equilibrium. If $\frac{1}{n^\gamma} \leq \epsilon$ we are done, so we assume n is smaller. We set $n' = (1/\epsilon)^{1/\gamma}$, so that $\frac{1}{n'^\gamma} = \epsilon$. Given an n -player anonymous game G , we build an n' -player anonymous game G' as follows: we add $n' - n$ dummy players, and give the dummy players utility 1 on strategy 1, and 0 on any other strategies so in any ϵ -well-supported equilibria, the dummy player must all play strategy 1 with probability 1. (Note that this is only true for ϵ -well-supported Nash equilibrium; in an ϵ -approximate Nash equilibrium, the dummy players can put ϵ probability elsewhere.) We shift the utility function of the actual players to ignore the dummy players on strategy 1. Formally, the payoff structure of G' is given by:

- For each $i > n$,

$$u_a^i(x) = \begin{cases} 1 & \text{if } a = 1 \\ 0 & \text{otherwise} \end{cases}$$

- For each $i \leq n$, we subtract the number of players on strategy 1 by $n' - n$ and then apply the original utility function. We define $\phi : \mathbb{Z}^k \rightarrow \mathbb{Z}^k$ as

$$\phi(x_1, \dots, x_k) = (x_1 - (n' - n), x_2, \dots, x_k),$$

$$u_a^i(x) = \begin{cases} u_a^i(\phi(x)) & \text{if } x_1 \geq n' - n \\ 0 & \text{otherwise} \end{cases}$$

Since $\epsilon = \frac{1}{n'^\gamma}$, by assumption we can compute an ϵ -well-supported equilibrium of G' in time $O(n'^d)$, and we can simply remove the dummy players to obtain an ϵ -equilibrium of the original game G . The running time is $O(n'^d) = \text{poly}(n, 1/\epsilon)$ when $\gamma = \Theta(1)$.

Proof. [Proof of Theorem 1.3] Assume that we can compute an $O(n^{-c})$ -approximate equilibrium in polynomial time for some constant $c > 1$. Let $\gamma = c - 1$, so we can compute an $O(\frac{1}{n^{1+\gamma}})$ -approximate equilibrium in polynomial time. By Lemma 4.1, we can convert it into an $O(\frac{1}{n^{\gamma/2}})$ -well-supported equilibrium. Lemma 4.2 then states that any polynomial-time algorithm that computes a well-supported Nash equilibrium of an inverse polynomial precision gives an FPTAS for computing well-supported Nash equilibria in anonymous games.

5 Proof of Theorem 1.2

In this section, we present a faster algorithm that computes an $\tilde{O}(n^{-1/3}k^{11/3})$ -approximate Nash equilibrium in n player k strategy anonymous games. Note that this algorithm always runs in polynomial time in the input size, without assuming any relationship between n and k .

Our approach builds on the idea of [GT15] to “smooth” an anonymous game by forcing all the players to randomize. We prove that the perturbed game is Lipschitz and therefore admits a pure Nash equilibrium (Lemma 5.1), which corresponds to simple approximate equilibria of a specific form in the original game: Each player plays one strategy with probability $1 - \delta$ for some small δ , and plays other strategies uniformly at random with probability δ . To prove the perturbed game is Lipschitz (Proposition 5.1), we rely on the recently established multivariate central limit theorem (CLT) of [DDKT16, DKS16a] to show that for $\delta = \Omega(n^{-1/3})$ the associated PMD is close to a discrete Gaussian.

Recall that an anonymous game $G = (n, k, \{u_a^i\}_{i \in [n], a \in [k]})$ is λ -Lipschitz if

$$|u_a^i(x) - u_a^i(y)| \leq \lambda \|x - y\|_1, \\ \forall i \in [n], \forall a \in [k], \quad \forall x, y \in \Pi_{n-1}^k.$$

An approximate pure Nash equilibrium always exists in Lipschitz anonymous games.

LEMMA 5.1. ([DP15, AS13]) *Every λ -Lipschitz anonymous game with k strategies admits a $(2k\lambda)$ -approximate pure Nash equilibrium. Moreover, such an*

approximate equilibrium can be found in time $\tilde{O}(n+k)$ times the description size of the game.

We perturb the input game G to get another game G_δ as follows. Let $X_\delta(e_j)$ denote the k -CRV that takes value e_j with probability $1-\delta$, and takes value $e_{j'}$ with probability $\frac{\delta}{k-1}$ for all other $j' \neq j$. When a player plays the strategy j in the perturbed game G_δ , it is as if she is playing $X_\delta(e_j)$ in the original game G . For example, the strategy $(1, 0, \dots, 0)$ in G_δ maps back to the mixed strategy $(1-\delta, \frac{\delta}{k-1}, \dots, \frac{\delta}{k-1})$ in G .

By forcing all players to randomize, we increase the uncertainty in the outcome of the game (i.e., the variance of the resulting PMD), and thus making the game “smoother”. As we will prove later, the perturbed game G_δ is λ -Lipschitz for $\lambda = \tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right)$. It then follows from Lemma 5.1 that there exists a $(2k\lambda)$ -pure Nash equilibrium of G_δ , which is a $(\delta + 2k\lambda)$ -mixed Nash equilibrium of G . The next proposition formally defines the payoff structure of G_δ , and bounds its Lipschitz constant.

PROPOSITION 5.1. *Given an anonymous game $G = (n, k, \{u_a^i\}_{i \in [n], a \in [k]})$ with payoffs normalized to $[0, 1]$, we define an anonymous game $G_\delta = (n, k, \{u_a^i\}_{i \in [n], a \in [k]})$ as follows,*

$$\begin{aligned} \forall i \in [n], a \in [k], x \in \Pi_{n-1}^k, \\ u_a^i(x) \stackrel{\text{def}}{=} (1-\delta) \mathbb{E}_{x' \sim M_\delta(x)} [u_a^i(x')] \\ + \frac{\delta}{k-1} \sum_{a' \neq a} \mathbb{E}_{x' \sim M_\delta(x)} [u_{a'}^i(x')], \end{aligned}$$

where $M_\delta(x) = \sum_{j \in [k]} x_j X_\delta(e_j)$ is an $(n-1, k)$ -PMD that corresponds to the perturbed outcome of the partition $x \in \Pi_{n-1}^k$. Then G_δ is $\tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right)$ -Lipschitz.

We defer the proof of Proposition 5.1 to the next subsection. We now show how Theorem 1.2 follows from Proposition 5.1.

Proof. [Proof of Theorem 1.2] Proposition 5.1 shows that G_δ is $\tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right)$ -Lipschitz. By Lemma 5.1, there exists a $(2k\lambda)$ -approximate pure Nash equilibrium in G_δ , and as noted in [DP15], such an approximate equilibrium can be found in total number of bit operations that is $\tilde{O}(n+k)$ times the description size of G_δ , by enumerating pure strategy profiles and solving maximum flows to match players to mixed strategies. Since we can compute the payoff structure of G_δ in polynomial-time given the input game G , the overall running time is polynomial in the input size.

We now bound the quality of the approximate Nash equilibrium. Note that a $(2k\lambda)$ -pure equilibrium of G_δ is a $(\delta + 2k\lambda)$ -mixed Nash equilibrium of G , since an ϵ -equilibrium in G_δ means that players cannot gain more than ϵ by deviating to the mixed strategies of the form $X_\delta(e_j) = (1-\delta)e_j + \frac{\delta}{k-1}(\mathbf{1} - e_j)$, so they gain at most $(\delta + 2k\lambda)$ by deviating to any e_j . Because changing what a player is doing δ fraction of the time can change her payoff by at most δ . Therefore, we can compute an $(\delta + 2k\lambda) = \tilde{O}\left(\delta + \frac{k^{11/2}}{\sqrt{n\delta}}\right)$ -equilibrium of the original game G in polynomial-time for any $\delta > 0$. Finally, setting $\delta = \frac{k^{11/3}}{n^{1/3}}$, we get an $\tilde{O}\left(\frac{k^{11/3}}{n^{1/3}}\right)$ -approximate Nash equilibrium.

5.1 Proof of Proposition 5.1 This section is devoted to the proof of Proposition 5.1. We will make use of the following two results. The first lemma is the multivariate central limit theorem from [DKS16a], which states that if an (n, k) -PMD X has high variance in all directions orthogonal to the all ones vector $\mathbf{1}$ (its variance along $\mathbf{1}$ is 0), then the projection of X on the first $(k-1)$ coordinates is close to a discretized Gaussian distribution with the same mean vector and covariance matrix.

LEMMA 5.2. ([DKS16A]) *Let X be an (n, k) -PMD, and X' be a $(k-1)$ -dimensional random variable that is the projection of X onto its first $k-1$ coordinates. Let Σ' be the covariance matrix of X' . Suppose that Σ' has no eigenvectors with eigenvalue less than σ'^2 . Let G' be the distribution obtained by sampling from $\mathcal{N}(\mathbb{E}[X'], \Sigma')$ and rounding to the nearest point in \mathbb{Z}^k . Then, we have that*

$$d_{\text{TV}}(X', G') \leq O\left(k^{7/2} \sqrt{\log^3(\sigma')/\sigma'}\right).$$

The second simple lemma states that if two k -dimensional Gaussian distributions have similar mean vectors and variances (in all directions), then they are close in total variation distance.

LEMMA 5.3. ([DDKT16]) *For two k -dimensional Gaussians $X \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $Y \sim \mathcal{N}(\mu_2, \Sigma_2)$, such that for all unit vector v ,*

$$|v^T(\mu_1 - \mu_2)| \leq \epsilon s_v, \text{ and } |v^T(\Sigma_1 - \Sigma_2)v| \leq \frac{\epsilon s_v^2}{2\sqrt{k}},$$

where $s_v^2 = \max\{v^T \Sigma_1 v, v^T \Sigma_2 v\}$. Then $d_{\text{TV}}(X, Y) \leq \epsilon$.

Proof. [Proof of Proposition 5.1] To prove the game G_δ

is λ -Lipschitz, we need to show that

$$\forall i \in [n], \forall a \in [k], \quad \forall x, y \in \Pi_{n-1}^k,$$

$$\left| \mathbb{E}_{x' \sim M_\delta(x)}[u_a^i(x')] - \mathbb{E}_{y' \sim M_\delta(y)}[u_a^i(y')] \right| \leq \lambda \|x - y\|_1.$$

In fact, because the payoff entries are normalized in $[0, 1]$, it is sufficient to show that the total variation distance between the $(n-1, k)$ -PMDs $M_\delta(x)$ and $M_\delta(y)$ is small, namely

$$\forall x, y \in \Pi_{n-1}^k, \quad d_{\text{TV}}(M_\delta(x), M_\delta(y)) \leq \lambda \|x - y\|_1.$$

Let $M'_\delta(x)$ and $M'_\delta(y)$ be the distributions $M_\delta(x)$ and $M_\delta(y)$ projected onto their first $k-1$ coordinates. Note that since all coordinates must sum to n , the k^{th} coordinate is redundant and so $d_{\text{TV}}(M_\delta(x), M_\delta(y)) = d_{\text{TV}}(M'_\delta(x), M'_\delta(y))$. To show that $M'_\delta(x)$ and $M'_\delta(y)$ are close in total variation distance, we first prove that the covariance matrix of $M'_\delta(x)$ has high variance in all directions, which allows us to use the multivariate central limit theorem (Lemma 5.2) to conclude that both $M'_\delta(x)$ and $M'_\delta(y)$ are close to the (discretized) Gaussian distributions with the same mean vectors and covariance matrices respectively. We then bound from above the total variation distance between two high-variance k -dimensional Gaussian distributions whose mean vectors are essentially x and y .

Recall that $M_\delta(x)$ is the sum of $n-1$ independent k -CRVs, and let Σ_1 denote the covariance matrix of $M_\delta(x)$. For any unit vector $v \in \mathbb{R}^k$ that is orthogonal to the all-one vector, we have

$$\begin{aligned} & \text{Var}[v^T X_\delta(e_j)] \\ &= \mathbb{E}\left[\left(v^T X_\delta(e_j)\right)^2\right] - \left(\mathbb{E}[v^T X_\delta(e_j)]\right)^2 \\ &= (1-\delta)v_j^2 + \frac{\delta}{k-1} \sum_{j' \neq j} v_{j'}^2 \\ &\quad - \left((1-\delta)v_j + \frac{\delta}{k-1} \sum_{j' \neq j} v_{j'} \right)^2 \\ &= (1-\delta)v_j^2 + \frac{\delta}{k-1} (1-v_j^2) \\ &\quad - \left((1-\delta)v_j - \frac{\delta}{k-1} v_j \right)^2 \\ &\geq \frac{\delta}{k-1}, \end{aligned}$$

where we simplify the expression using the fact that $\sum_j v_j = 0$ and $\sum_j v_j^2 = 1$, and then take derivative to minimize it. Therefore, for any unit vector v , $v^T \Sigma_1 v = \text{Var}[v^T M_\delta(x)] = \sum_{j \in [k]} x_j \text{Var}[v^T X_\delta(e_j)] \geq \frac{(n-1)\delta}{k-1}$, which implies that Σ_1 has no eigenvalues less

than $\frac{(n-1)\delta}{k-1}$ (except the one associated with $\mathbf{1}$). We then use the following lemma to bound from below the eigenvalues of Σ'_1 :

LEMMA 5.4. *Suppose that Σ is a positive semidefinite matrix with $\Sigma \mathbf{1} = 0$ and that all other eigenvalues of Σ are at least σ^2 . Then for all vectors $w \in \mathbb{R}^k$ with $w_k = 0$, we have that*

$$\frac{w^T \Sigma w}{w^T w} \geq \sigma^2 / k.$$

Proof. Let w be a vector that minimizes $\frac{w^T \Sigma w}{w^T w}$ over $w \in \mathbb{R}^k$ with $w_k = 0$. Then $v = w - \frac{w^T \mathbf{1}}{k} \mathbf{1}$ has $v^T \mathbf{1} = 0$ and so $v^T \Sigma v \geq \sigma^2 v^T v$. We have $v^T \Sigma v = w^T \Sigma w$ since $v - w$ is a multiple of $\mathbf{1}$, and we have

$$\begin{aligned} v^T v &= \left(w - \frac{w^T \mathbf{1}}{k} \mathbf{1} \right)^T \left(w - \frac{w^T \mathbf{1}}{k} \mathbf{1} \right) \\ &= w^T w + (w^T \mathbf{1})^2 / k - 2(w^T \mathbf{1})^2 / k \\ &= \|w\|_2^2 - \|w\|_1^2 / k \\ &\geq w^T w / k, \end{aligned}$$

where the last line follows from the inequality $\|w\|_1 \leq \sqrt{k-1} \|w\|_2$. Thus, we have that

$$\frac{w^T \Sigma w}{w^T w} \geq \frac{v^T \Sigma v}{k v^T v} \geq \sigma^2 / k.$$

Since all except one eigenvalues of each of Σ_1 and Σ_2 are at least $\frac{(n-1)\delta}{(k-1)}$, the minimum eigenvalues of Σ'_1 and Σ'_2 are at least $\frac{(n-1)\delta}{k^2}$. Let $\mathcal{Z}(\mu, \Sigma)$ be the discretized Gaussian obtained by rounding $\mathcal{N}(\mu, \Sigma)$ to the nearest integer in every coordinate. Then, by Lemma 5.2, we have

$$\begin{aligned} (5.4) \quad d_{\text{TV}}(M'_\delta(x), \mathcal{Z}(\mu'_1, \Sigma'_1)) &\leq \tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right) \\ d_{\text{TV}}(M'_\delta(y), \mathcal{Z}(\mu'_2, \Sigma'_2)) &\leq \tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right). \end{aligned}$$

Next, we use Lemma 5.3 to bound the total variation distance between the k -dimensional Gaussian distributions $\mathcal{N}(\mu'_1, \Sigma'_1)$ and $\mathcal{N}(\mu'_2, \Sigma'_2)$. Let μ_1, μ_2 and Σ_1, Σ_2 be the mean vectors and the covariance matrices of $M_\delta(x)$ and $M_\delta(y)$ respectively. Observe that

$$\mu_1 = \left(1 - \frac{k\delta}{k-1}\right) x + \delta \mathbf{1}.$$

So, for any unit vector $v \in \mathbb{R}^k$,

$$\begin{aligned} s_v^2 &= \max\{v^T \Sigma_1 v, v^T \Sigma_2 v\} \geq \frac{(n-1)\delta}{k-1}, \\ |v^T(\mu_1 - \mu_2)| &= \left| v^T \left(\left(1 - \frac{k\delta}{k-1}\right) x - \left(1 - \frac{k\delta}{k-1}\right) y \right) \right| \\ &\leq \left(1 - \frac{k\delta}{k-1}\right) \|x - y\|_1 \leq \|x - y\|_1. \end{aligned}$$

If the unit vector v is orthogonal to $\mathbf{1}$, we can use the expression for $v^T \Sigma_1 v$ we had earlier. Taking derivative with respect to v_j shows that $\text{Var}[v^T X_\delta(e_j)]$ is maximized at $v = \pm e_j$. Hence, we can write

$$\begin{aligned} |v^T(\Sigma_1 - \Sigma_2)v| &= \sum_{j \in [k]} (x_j - y_j) \text{Var}[v^T X_\delta(e_j)] \\ &\leq \|x - y\|_1 \max_j \text{Var}[v^T X_\delta(e_j)] \\ &\leq \|x - y\|_1 \left[(1 - \delta) - \left(1 - \frac{k\delta}{k-1}\right)^2 \right] \\ &= \frac{k-1}{k+1} \delta \|x - y\|_1 \leq \delta \|x - y\|_1. \end{aligned}$$

To see that the upper bound on $|v^T(\Sigma_1 - \Sigma_2)v|$ holds for all unit vectors, observe that for both covariance matrices it holds $\Sigma_1 \mathbf{1} = \Sigma_2 \mathbf{1} = \mathbf{0}$. For any unit vector v' , we can take its projection onto the subspace orthogonal to $\mathbf{1}$, and write v' as a linear combination $\alpha v + \beta \mathbf{1}$, for some $\alpha < 1$ and a unit vector v that is orthogonal to $\mathbf{1}$. That is,

$$\begin{aligned} |v'^T(\Sigma_1 - \Sigma_2)v'| &= |(\alpha v + \beta \mathbf{1})^T(\Sigma_1 - \Sigma_2)(\alpha v + \beta \mathbf{1})| \\ &= |\alpha^2 v^T(\Sigma_1 - \Sigma_2)v| \leq |v^T(\Sigma_1 - \Sigma_2)v|. \end{aligned}$$

Thus, for all unit vectors $v \in \mathbb{R}^k$, we have $|v^T(\mu_1 - \mu_2)| \leq \|x - y\|_1$ and $|v^T(\Sigma_1 - \Sigma_2)v| \leq \delta \|x - y\|_1$. In particular, this holds for vectors with k^{th} coordinate 0. Hence, for all $v \in \mathbb{R}^{k-1}$, we have $|v^T(\mu'_1 - \mu'_2)| \leq \|x - y\|_1$ and $|v^T(\Sigma'_1 - \Sigma'_2)v| \leq \delta \|x - y\|_1$.

Finally, we set $\epsilon = O\left(\frac{\sqrt{k}}{\sqrt{n\delta}} + \frac{k^{3/2}}{n}\right) \|x - y\|_1$ to satisfy the requirements of Lemma 5.3, and therefore $d_{\text{TV}}(\mathcal{N}(\mu'_1, \Sigma'_1), \mathcal{N}(\mu'_2, \Sigma'_2)) \leq \epsilon$. By the data processing inequality, rounding both distributions to the nearest integer coordinates does not increase their total variation distance, therefore

$$(5.5) \quad d_{\text{TV}}(\mathcal{Z}(\mu'_1, \Sigma'_1), \mathcal{Z}(\mu'_2, \Sigma'_2)) \leq \epsilon.$$

By the triangle inequality, Equations (5.4) and (5.5)

yield

$$\begin{aligned} d_{\text{TV}}(M_\delta(x), M_\delta(y)) &= d_{\text{TV}}(M'_\delta(x), M'_\delta(y)) \\ &\leq d_{\text{TV}}(M'_\delta(x), \mathcal{Z}(\mu'_1, \Sigma'_1)) \\ &\quad + d_{\text{TV}}(\mathcal{Z}(\mu'_1, \Sigma'_1), \mathcal{Z}(\mu'_2, \Sigma'_2)) \\ &\quad + d_{\text{TV}}(\mathcal{Z}(\mu'_2, \Sigma'_2), M'_\delta(y)) \\ &\leq \tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right) + O\left(\frac{\sqrt{k}}{\sqrt{n\delta}} + \frac{k^{3/2}}{n}\right) \|x - y\|_1 \\ &\leq \tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right) \|x - y\|_1. \end{aligned}$$

The last inequality holds because $M_\delta(x) = M_\delta(y)$ when $x = y$, so we can assume that $\|x - y\|_1 \geq 1$. This concludes the proof that G_δ is λ -Lipschitz for $\lambda = \tilde{O}\left(\frac{k^{9/2}}{\sqrt{n\delta}}\right)$.

References

- [AS13] Yaron Azrieli and Eran Shmaya. Lipschitz games. *Mathematics of Operations Research*, 38(2):350–357, 2013.
- [Blo99] Matthias Blonski. Anonymous games with binary actions. *Games and Economic Behavior*, 28(2):171–180, 1999.
- [Blo05] Matthias Blonski. The women of Cairo: Equilibria in large anonymous games. *Journal of Mathematical Economics*, 41(3):253–264, 2005.
- [CDO15] Xi Chen, David Durfee, and Anthi Orfanou. On the complexity of Nash equilibria in anonymous games. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 381–390, 2015.
- [DDKT16] Constantinos Daskalakis, Anindya De, Gautam Kamath, and Christos Tzamos. A size-free CLT for Poisson multinomials and its applications. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1074–1086, 2016.
- [DDS12] Constantinos Daskalakis, Ilias Diakonikolas, and Rocco A. Servedio. Learning poisson binomial distributions. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 709–728. ACM, 2012.
- [DGP09] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM J. Comput.*, 39(1):195–259, 2009.
- [DKS16a] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. The Fourier transform of Poisson multinomial distributions and its algorithmic applications. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1060–1073, 2016.

- [DKS16b] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Optimal learning via the fourier transform for sums of independent integer random variables. In *Proceedings of The 29th Conference on Learning Theory (COLT)*, pages 831–849, 2016.
- [DKS16c] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Properly learning Poisson binomial distributions in almost polynomial time. In *Proceedings of The 29th Conference on Learning Theory (COLT)*, pages 850–878, 2016.
- [DP07] Constantinos Daskalakis and Christos H. Papadimitriou. Computing equilibria in anonymous games. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 83–93, 2007.
- [DP08] Constantinos Daskalakis and Christos H. Papadimitriou. Discretized multinomial distributions and Nash equilibria in anonymous games. In *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 25–34, 2008.
- [DP09] Constantinos Daskalakis and Christos H. Papadimitriou. On oblivious PTAS’s for Nash equilibrium. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 75–84, 2009.
- [DP15] Constantinos Daskalakis and Christos H. Papadimitriou. Approximate Nash equilibria in anonymous games. *J. Economic Theory*, 156:207–245, 2015.
- [GT15] Paul W. Goldberg and Stefano Turchetta. Query complexity of approximate equilibria in anonymous games. In *Proceedings of the 11th Conference on Web and Internet Economics (WINE)*, pages 357–369, 2015.
- [LMM03] Richard J. Lipton, Evangelos Markakis, and Aranyak Mehta. Playing large games using simple strategies. In *Proceedings 4th ACM Conference on Electronic Commerce (EC)*, pages 36–41, 2003.
- [Mil96] Igal Milchtaich. Congestion games with player-specific payoff functions. *Games and Economic Behavior*, 13(1):111–124, 1996.
- [Rub16] Aviad Rubinstein. Settling the complexity of computing approximate two-player Nash equilibria. In *57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2016.
- [Sim82] Herbert A. Simon. *Models of bounded rationality: Empirically grounded economic reason*, volume 3. MIT press, 1982.

Appendix

A Proof of Lemma 3.7

This lemma is a generalization of Lemma 5.5 of [DKS16a], which assumes that $\epsilon = \widetilde{O}_k(1/\sigma)$. Thus, we need to be careful about where this relation was used in the proof.

Note that for a fixed ξ , if ξ' satisfies $\xi' \in \xi + \mathbb{Z}^k$ and $\xi'^T \Sigma \xi' \leq Ck \log(1/\epsilon)$, then so does $\xi + i\mathbf{1}$ for all $i \in \mathbb{Z}$. We define T' as $T' \stackrel{\text{def}}{=} \{ \xi' \in \mathbb{R}^k : \xi'^T \Sigma \xi' \leq Ck \log(1/\epsilon) \text{ and } 0 \leq \xi' \cdot \mathbf{1} \leq k \}$. Then, $\xi \in T$ if and only if there is a $\xi' \in T'$ with

$$\xi - \xi' \in \mathbb{Z}^k.$$

- (i) Because $\xi - \xi' \in \mathbb{Z}^k$, we have $|\xi_i - \xi_j| \leq |\xi'_i - \xi'_j|$. So to prove (i), we need to show that $|\xi'_i - \xi'_j| \leq 2\sqrt{Ck \log(1/\epsilon)}/\sigma$ for all $\xi' \in T'$, i and j .

Fix $\xi' \in T'$, we define $\tilde{\xi}'$ to be the projection of ξ' onto the subspace orthogonal to $\mathbf{1}$, i.e., $\tilde{\xi}' = \xi' - \frac{\xi' \cdot \mathbf{1}}{k} \mathbf{1}$. Since $\Sigma \mathbf{1} = \mathbf{0}$ and all other eigenvalues of Σ are at least σ^2 , for all i, j we have

$$\begin{aligned} |\xi'_i - \xi'_j| &= |\tilde{\xi}'_i - \tilde{\xi}'_j| \leq 2\|\tilde{\xi}'\|_\infty \leq 2\|\tilde{\xi}'\|_2 \\ &\leq 2\sqrt{\xi'^T \Sigma \xi' / \sigma^2} \leq 2\sqrt{Ck \log(1/\epsilon)}/\sigma. \end{aligned}$$

This proves (i).

- (ii) Next we consider $\text{Vol}(T')$. If $\xi' \in T'$, we know that $\|\xi' - (\xi' \cdot \mathbf{1}/k)\mathbf{1}\|_2^2 \leq Ck \log(1/\epsilon)/\sigma^2$. Also $0 \leq \xi' \cdot \mathbf{1} \leq k$ implies that $\|(\xi' \cdot \mathbf{1}/k)\mathbf{1}\|_2^2 \leq k$. Because these two vectors are orthogonal, we can write

$$\begin{aligned} \|\xi'\|_2^2 &= \left\| \xi' - \frac{\xi' \cdot \mathbf{1}}{k} \mathbf{1} \right\|_2^2 + \left\| \frac{\xi' \cdot \mathbf{1}}{k} \mathbf{1} \right\|_2^2 \\ &\leq Ck \log(1/\epsilon)/\sigma^2 + k \leq 2Ck \log(1/\epsilon), \end{aligned}$$

where the last inequality holds by the assumption that $\sigma \geq 1$. Thus,

$$\xi'^T (\Sigma + I) \xi' = \xi'^T \Sigma \xi' + \|\xi'\|_2^2 \leq 3Ck \log(1/\epsilon).$$

By Claim 5.4 of [DKS16a], we get that

$$\text{Vol}(T') \leq \det(\Sigma + I)^{-1/2} O(C \log(1/\epsilon))^{k/2}.$$

It then follows from Lemma 3.6 that $\text{Vol}(T')|S| = O(C \log(1/\epsilon))^k$.

To show (ii), we need to show that $\text{Vol}(T) \leq \text{Vol}(T')$. Note that T' is a disjoint union of its intersections with unit cubes with integer corners, and so

$$\text{Vol}(T') = \sum_{b \in \mathbb{Z}^k} \text{Vol} \left(T' \cap \prod_{i=1}^k [b_i, b_i + 1) \right).$$

On the other hand, T is the union of translations of these sets

$$T = \bigcup_{b \in \mathbb{Z}^k} \{ \xi' - b : \xi' \in T' \cap \prod_{i=1}^k [b_i, b_i + 1) \},$$

so $\text{Vol}(T) \leq \text{Vol}(T')$.

(iii) By the pigeonhole principle, for every $\xi \in \mathbb{R}^k$, there is an interval I_ξ of length $\frac{k}{k+1}$ such that there exists $\xi' \in \xi + \mathbb{Z}^k$ where all the coordinates of ξ' are in I_ξ . We define T_m to be

$$T_m \stackrel{\text{def}}{=} \left\{ \xi \in [0, 1]^k : \exists \xi' \in (\xi + \mathbb{Z}^k) \cap I_\xi^k \text{ and } 2^m C k \log(\sigma) \leq \xi'^T \Sigma \xi' \leq 2^{m+1} C k \log(\sigma) \right\}.$$

Then, we have that $T \cup (\bigcup_{m=0}^\infty T_m) = [0, 1]^k$, although these sets need not be disjoint. Thus, $[0, 1]^k / T \subseteq \bigcup_{m=0}^\infty T_m$ and so

$$\int_{[0, 1]^k / T} |\hat{X}(\xi)| d\xi \leq \sum_{m=0}^\infty \text{Vol}(T_m) \sup_{\xi \in T_m} |\hat{X}(\xi)|.$$

If we apply (ii) of this lemma with $2^{m+1}C$ instead of C , the resulting set T would be a superset of T_m . Thus, we have that $\text{Vol}(T_m) \leq O(2^{m+1}C \log(1/\epsilon))^k / |S|$. To show (iii), we bound $\sup_{\xi \in T_m} |\hat{X}(\xi)|$ using the following claim, which gives a ‘‘Gaussian decay’’ upper bound on the magnitude of the Fourier transform.

CLAIM A.1. *For $\xi \in T_m$, it holds $|\hat{X}(\xi)| \leq \exp(-\Omega(C2^m \log(1/\epsilon)/k))$. If additionally we have $m \leq 3 \log_2 k$, then $|\hat{X}(\xi)| = \exp(-\Omega(C2^m k \log(1/\epsilon)))$.*

Proof. We take $\xi' \in (\xi + \mathbb{Z}^k) \cap I_\xi^k$ as in the definition of T_m . Lemma 3.10 of [DKS16a] gives that if the coordinates of ξ' lie in an interval of length $1 - \delta$, then

$$\begin{aligned} |\hat{X}(\xi)| &= |\hat{X}(\xi')| \leq \exp(-\Omega(\delta^2 \xi'^T \cdot \Sigma \cdot \xi')) \\ &= \exp(-\Omega(C2^m k \log(1/\epsilon) \delta^2)). \end{aligned}$$

By the definition of T_m , we take $\delta = \frac{1}{k+1}$ to get the bound $|\hat{X}(\xi)| \leq \exp(-\Omega(C2^m \log(1/\epsilon)/k))$.

To get the stronger bound, we need to show that when m is small all coordinates of ξ' are in a shorter interval. This is because, if we apply (i) of this lemma with $2^{m+1}C$ instead of C , we have $|\xi'_i - \xi'_j| \leq \sqrt{2^{m+3} C k \log(1/\epsilon) / \sigma}$ for any i, j . When $m \leq \log_2(\sigma / (C k \log(1/\epsilon))) - 4$, we can take $\delta = 1/2$ and obtain the stronger bound of the claim.

This is where we use our assumption that $\sigma \geq \text{poly}(k \log(1/\epsilon))$. We need $m \leq 3 \log_2 k \leq \log_2(\sigma / (C k \log(1/\epsilon))) - 4$, which holds when $\sigma \geq 16 C k^4 \log(1/\epsilon)$.

Finally, for (iii) we can write

$$\begin{aligned} \int_{[0, 1]^k / T} |\hat{X}(\xi)| d\xi &\leq \sum_{m=0}^\infty \text{Vol}(T_m) \sup_{\xi \in T_m} |\hat{X}(\xi)| \\ &\leq \sum_{m=0}^\infty O(2^{m+1} C \log(1/\epsilon))^k \sup_{\xi \in T_m} |\hat{X}(\xi)| \\ &\leq \frac{O(C \log(1/\epsilon))^k}{|S|} \sum_{m=0}^\infty 2^{mk} \sup_{\xi \in T_m} |\hat{X}(\xi)|. \end{aligned}$$

We divide this sum into two pieces:

$$\begin{aligned} \sum_{m=0}^{3 \log_2 k} 2^{mk} \sup_{\xi \in T_m} |\hat{X}(\xi)| &\leq \sum_{m=0}^{3 \log_2 k} 2^{mk} \exp(-\Omega(C2^m k \log(1/\epsilon))) \\ &\leq \sum_{m=0}^{3 \log_2 k} \exp(-\Omega(C(2^m - m)k \log(1/\epsilon))) \\ &\leq \sum_{m=0}^{3 \log_2 k} 2^{-m} \exp(-\Omega(Ck \log(1/\epsilon))) \\ &\leq \exp(-\Omega(Ck \log(1/\epsilon))) = \epsilon^{\Omega(Ck)} \end{aligned}$$

and

$$\begin{aligned} \sum_{m=3 \log_2 k}^\infty 2^{mk} \sup_{\xi \in T_m} |\hat{X}(\xi)| &\leq \sum_{m=3 \log_2 k}^\infty 2^{mk} \exp(-\Omega(C2^m \log(1/\epsilon)/k)) \\ &\leq \sum_{m=3 \log_2 k}^\infty \exp(-\Omega(C(2^m - mk^2) \log(1/\epsilon)/k)) \\ &\leq \sum_{m=3 \log_2 k}^\infty \exp(-\Omega(C(k^2 + mk) \log(1/\epsilon)/k)) \\ &\leq \sum_{m=3 \log_2 k}^\infty 2^{-m} \exp(-\Omega(Ck \log(1/\epsilon))) \leq \epsilon^{\Omega(Ck)}. \end{aligned}$$

We thus have

$$\begin{aligned} \int_{[0, 1]^k \setminus T} |\hat{X}(\xi)| d\xi &\leq O(C \log(1/\epsilon))^k \epsilon^{\Omega(Ck)} / |S| \\ &\leq \epsilon / (2|S|). \end{aligned}$$